

WE CLAIM:

1. A method for merging a source document into a destination document to create a merged document, wherein the source document or the destination document or both contain one or more markup language tags having a start tag or an end tag or both a start tag and an end tag, the method comprising:

comparing the source document and the destination document to identify one or more matching blocks and one or more difference blocks;

splitting all matching blocks that contain a markup language tag for which only a start tag or an end tag has been matched so that the matched markup language tags are no longer matched;

splitting all matching blocks that contain a markup language tag for which either the start tag or the end tag has been matched to a different tag in the other document so that the matched markup language tags are no longer matched;

splitting any matching blocks containing markup language tags that would create overlapping tags when merged; and

merging the matching and difference blocks so that all markup language tags contained in the merged document are well formed.

2. The method of Claim 1, wherein splitting any matching blocks containing markup language tags that would create overlapping tags when merged comprises for each difference block:

determining whether a markup language tag from the source document has a start tag but not an end tag in the difference block;

in response to determining that the difference block contains a start tag but not an end tag, locating the furthest difference block containing the matching end tag of such a tag in the source document;

identifying an inner range of matching and difference blocks as each of the blocks between the current difference block and the furthest difference block;

identifying an outer range of matching and difference blocks as the current difference block and the furthest difference block and each of the matching and difference blocks in between;

identifying each markup language tag that has a start tag within the inner range and an end tag outside the outer range and, for each identified markup language tag, discarding all matching blocks within the inner range prior to the start tag; and

identifying each markup language tag that has an end tag within the inner range and a start tag outside the outer range and, for each identified markup language tag, discarding all matching blocks within the inner range after the end tag.

3. The method of Claim 2, wherein merging the matching and difference blocks so that all markup language tags contained in the merged document are well formed comprises:

processing each matching block and difference block in order, and  
for each matching block, copying the text from the source or destination document in the matching block into the merged document; and

for each difference block:

copying the text from the source or destination document in the difference block into the merged document if the difference block contains text only from either the source or destination document.

4. The method of Claim 3, wherein merging the matching and difference blocks so that all markup language tags contained in the merged document are well formed further comprises, for each difference block:

creating a label for each markup language tag that has only either a start tag or an end tag in the difference block, the label comprising a non-zero integer identifying the number of blocks to the difference block containing the matching start or end tag;

creating a sub-block for each distinct label, each sub-block containing the markup language tags and text from either the source or destination document having a corresponding label;

arranging in order each of the sub-blocks, with negative labels first and then positive labels in decreasing order; and

copying text into the merged document from each sub-block, text from the source document being copied just prior to or just after text from the sub-block in the destination document.

5. The method of Claim 4, wherein text is copied from each sub-blocks into the merged document in the arranged order.

6. The method of Claim 5, wherein text in the sub-block from the source document is copied just prior to text from the sub-block from the destination document if the sub-block has a negative label.

7. The method of Claim 6, wherein text in the sub-block from the source document is copied just after text in the sub-block from the destination document if the sub-block has a positive label.

8. The method of Claim 7, wherein comparing the source document and the destination document to identify one or more matching blocks and one or more difference blocks comprises:

matching a markup language tag in the source document to a markup language tag in the destination document only if each aspect of the markup language tag in the source document is identical to each aspect of the markup language tag in the destination document.

9. A computer-readable medium having computer-executable instructions stored thereon which, when executed by a computer, will cause the computer to perform the method of Claim 1.

10. A computer-controlled apparatus capable of performing the method of  
Claim 1.